Remy Lagrois

Experimental Statistics

Project 3

Cluster Analysis of Osteoporosis in Women Over 55

**Introduction**

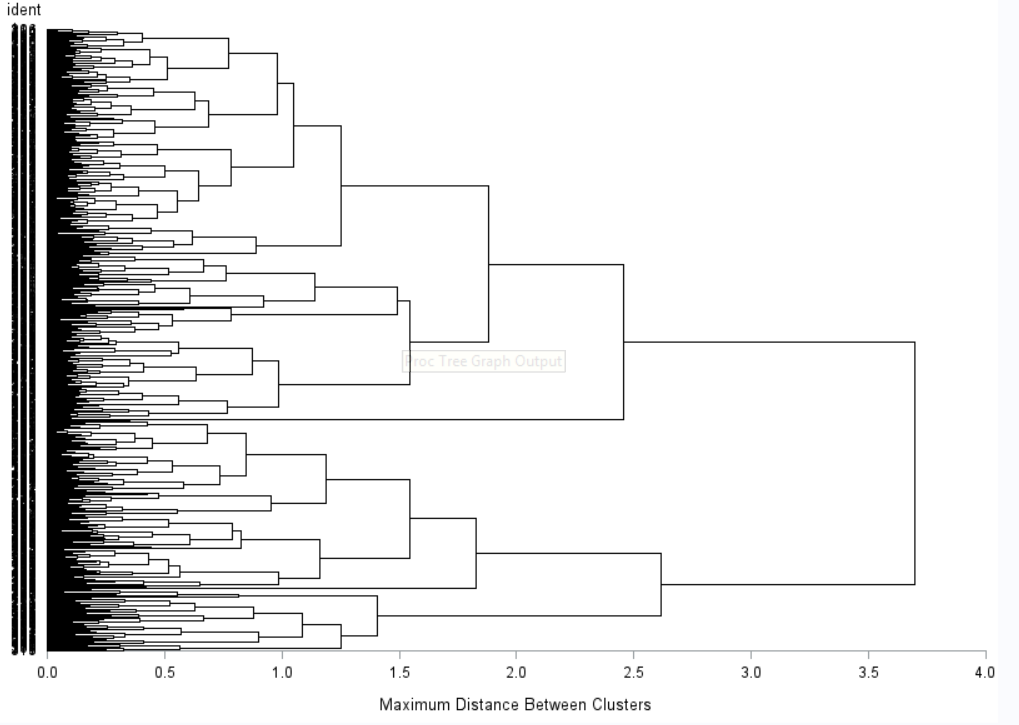
As any person gets older their chances of fracturing a bone increase as their bone density decreases (i.e. loss of calcium). This problem is especially apparent in women. As a woman moves out of reproductive age she will go through the process of menopause. While the primary effect of menopause is the halting of menstruation, it also includes an increased risk of osteoporosis.

The Global Longitudinal study of Osteoporosis in Women (GLOW) is, as the name suggests, a study looking at osteoporosis, risk factors, and outcomes in women over the age of 55. The primary goals are to identify the primary risk factors and reduce fractures. This is important both on the personal and national level. Breaking a bone is not a positive experience at any age but as a person gets older the consequences of a fracture (and the ability to recover) become worse and worse. This also can have economic impacts as the person or their family are prevented from working and causing an increase in insurance premiums more widely. Reducing bone fractures in the elderly will mean fewer people having to go through the experience and reduced costs (literally) for everyone else.

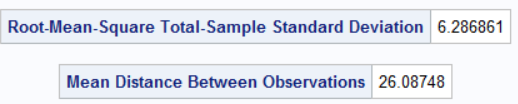
In this paper we perform a cluster analysis on five-hundred GLOW observations. We use complete linkage to try to find previously hidden groupings of women and their (potential) fractures. We then ran an analysis of variance on the clusters to find differences between them in terms of the means of our variables.

**The Data**

As previously mentioned we used five-hundred observations to run the analysis. The GLOW researchers collected the following information about each woman in the study: ID, study site, physician ID, history of fractures (hist), age of enrollment (age), weight at enrollment (wt), height at enrollment (ht), BMI (bmi), menopause before or after 45 (meno\_45), mother’s hip fractured or not (mom\_frac), use of arms needed when standing or not (arm\_ast), smoker or not, self-reported risk (risk), risk score (frac\_score), and any fractures in the past year (frac). For our cluster analysis we did not include any of the IDs as they don’t have anything to do with risk of fracture and we also did not include smoking information since there were so few smokers in the study and its removal improved our clusters.

**Analysis** 

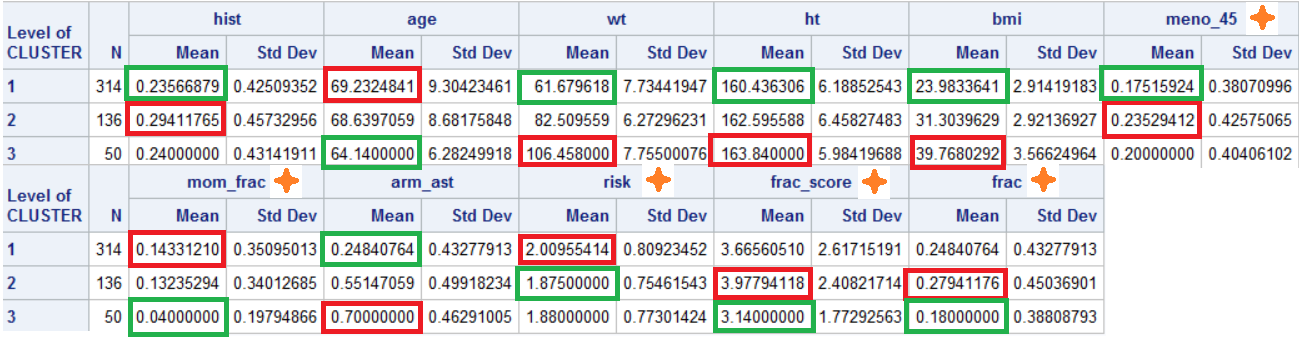
Above is the dendrogram resulting from running a cluster analysis using complete linkage and including the variables as described above. As there are 500 observations it is difficult to pick out where each individual is. However, we can see that the data is split into two main groupings at a distance between 3.5 and 4. That first split is somewhat uneven with slightly more than 60% falling into the top group.



The standard deviation of the total-sample root-mean-square came out to 6.29 while the mean distance is 26.1. Using the other linking algorithms resulted in standard deviations of up to around 8 and mean distances of up to around 29.

After running proc cluster and proc tree were run an analysis of variance was run using proc glm. This was done to find the mean of each attribute inside each cluster and to find if the differences in means were statistically significant. The value of ‘nclusters’ was set to three. Specifying a higher number resulted in several clusters containing only one or two observations while selecting fewer resulted in indistinct clusters while retaining clusters of one or two observations.

Of the eleven attributes used in our analysis, six were found to have statistically significant differences between the clusters (i.e. p-value below 0.05). See the appendix for the ANOVA tables. Those six significant attributes were: history of fractures, age, weight, height, BMI, and if they need to use their arms to help them stand up. The insignificant (p-value about 0.05) ones were: menopause before 45, whether or not their mother fractured her hip, self-assessed risk score, the calculated risk score, and whether or not they had a fracture in the last year.



The above table lists the attribute means of each cluster. The red box represents the cluster with the maximum average and the green box shows the minimum. Any attribute marked with the orange diamond did not have significant differences in the mean between clusters. Cluster 1 and cluster 3 are relatively easy to analyze while cluster 2 is somewhat less clear. Cluster 1 has the lowest average weight and BMI. The lowest proportion of people who have a history of fractures as well as the lowest proportion of people who need their arms when standing also belong to cluster 1. Cluster 1 clearly is women who are of a healthy body weight and no real history of bone fractures. Cluster 3 is basically the opposite. It has the heaviest women with the largest BMI. Cluster 3 also has the highest proportion of people who use their arms when standing. This would indicate cluster 3 contains women who are overweight and out of shape. Most of the means in cluster 2 are right between the ones in clusters 1 and 3. Cluster 2 does contain five minimums or maximums but four of those are in attributes that do not have a significant difference between the clusters. That one significant attribute is the history of fractures; cluster 2 has the highest proportion of people with such a history. Even though we cannot be sure they are actually different, cluster 2 also has the highest proportion of people who entered menopause before 45 and who had a fracture in the last year. It also has the highest average calculated risk score. This would indicate this cluster could be people of intermediate to somewhat overweight who are otherwise prone to fractures.

|  |  |
| --- | --- |
| Cluster | Meaning |
| One | Healthy Weight |
| Two | Fracture Prone |
| Three | Overweight |

**Conclusion**

The Global Longitudinal study of Osteoporosis in Women is an ongoing effort to study the risks and causes of osteoporosis in women over the age of 55. This analysis was done to analyze how 500 of the studied women could be grouped up based on the attributes that are being collected. We found there are three basic groupings: women of healthy weight, women who are fracture prone, and women who are overweight. These groupings would suggest that a person’s weight plays a role in bone fractures. This would stand to reason, more weight results in greater stress on bones. Further research should be done to see if increased weight does actually increase the risk and rate of fractures and how it affects the severity of osteoporosis.

**Appendix**

*Code*

**data** ost;

infile '\\Client\D$\SMU\Stats II\Project 3\glow500.csv' firstobs = **2** dlm = ',';

input sub\_id site\_id phy\_id hist age wt ht bmi meno\_45 mom\_frac arm\_ast smoke risk frac\_score frac;

ident = \_n\_;

drop sub\_id site\_id phy\_id smoke;

**run**;

**proc** **sort**;

by ident;

**run**;

title 'comp recheck';

**proc** **cluster** method=complete outtree=clust1;

var hist age wt ht bmi meno\_45 mom\_frac arm\_ast risk frac\_score frac;

id ident;

**run**;

**proc** **tree** horizontal nclusters=**3** out=clust2;

id ident;

**run**;

**proc** **sort**;

by ident;

**run**;

/\*

proc print;

run;

\*/

**data** combine;

merge ost clust2;

by ident;

**run**;

**proc** **glm** data=combine;

class cluster;

model hist age wt ht bmi meno\_45 mom\_frac arm\_ast risk frac\_score frac = cluster;

means cluster;

**run**;

*Misc Tables*

